

On the DMT of TDD-SIMO Systems with Channel-Dependent Reverse Channel Training

B. N. Bharath and Chandra R. Murthy *Senior Member, IEEE*

Abstract—This paper investigates the Diversity-Multiplexing gain Trade-off (DMT) of a training based reciprocal Single Input Multiple Output (SIMO) system, with (i) perfect Channel State Information (CSI) at the Receiver (CSIR) and noisy CSI at the Transmitter (CSIT), and (ii) noisy CSIR and noisy CSIT. In both the cases, the CSIT is acquired through Reverse Channel Training (RCT), i.e., by sending a training sequence from the receiver to the transmitter. A channel-dependent fixed-power training scheme is proposed for acquiring CSIT, along with a forward-link data transmit power control scheme. With perfect CSIR, the proposed scheme is shown to achieve a diversity order that is quadratically increasing with the number of receive antennas. This is in contrast with conventional orthogonal RCT schemes, where the diversity order is known to saturate as the number of receive antennas is increased, for a given channel coherence time. Moreover, the proposed scheme can achieve a larger DMT compared to the orthogonal training scheme. With noisy CSIR and noisy CSIT, a three-way training scheme is proposed and its DMT performance is analyzed. It is shown that nearly the same diversity order is achievable as in the perfect CSIR case. The time-overhead in the training schemes is explicitly accounted for in this work, and the results show that the proposed channel-dependent RCT and data power control schemes offer a significant improvement in terms of the DMT, compared to channel-agnostic orthogonal RCT schemes. The outage performance of the proposed scheme is illustrated through Monte-Carlo simulations.

Index Terms—Diversity-multiplexing gain tradeoff, MMSE channel estimation, training sequence.

I. INTRODUCTION

Reliability and system throughput are two fundamental parameters of interest in any wireless communication system, and the inherent tradeoff between the two at high SNR was elegantly captured by the Diversity Multiplexing gain Tradeoff (DMT) proposed in the seminal work of Zheng and Tse [2]. It is known that a significant improvement in the outage performance can be obtained if the Channel State Information (CSI) at the receiver (CSIR) and the transmitter (CSIT) are perfect [3], [4], while [2] considered perfect CSIR and no CSIT.

In a Time Division Duplex (TDD) system, CSI could be estimated at the transmitter and receiver by sending a known training sequence in the forward and reverse-link directions, respectively. This has two consequences. First, the *estimation*

error results in incorrect data rate or power adaptation at the transmitter, in turn leading to higher outage rate. Second, training incurs a *time overhead*, which could be non-trivial when the training occupies a significant fraction of the channel coherence time, as it affects the pre-log term in the achievable data rate [5]. This paper therefore focuses on the important problem of analytically comparing the DMT performance of different channel estimation techniques and identifying training signals and data power control schemes that result in a good performance in terms of the achievable DMT. We start with a brief survey of related literature.

The impact of imperfect CSIT on the DMT of a multiple antenna system has been a popular area of research, and it is known that even with imperfect CSIR and CSIT, a significant improvement in DMT can be obtained, compared to the no-CSIT case (see, for example, [6]–[8]). The effect of imperfect CSIR on the DMT of a MIMO system was first studied in [9]. The DMT analysis of a multiple antenna system with perfect CSIR and when the CSIT is modeled as the CSI plus Gaussian noise whose variance decreases with training SNR was investigated in [10]–[12]. In a TDD setup, the achievable DMT improvement using power control based on noisy CSIT was shown in [12]–[14]. Other works that study the DMT performance with quantized feedback of CSI and/or target data rate control based on noisy CSIT include [6], [7], [11], [15]–[18]. In [17], [19], the DMT of two-way and multi-round training schemes in a TDD system was derived. In these studies, the channel feedback signal on the reverse link is chosen to satisfy an average power constraint, rather than an instantaneous power constraint.

Most of the aforementioned studies of the DMT with imperfect CSI typically ignore the training duration overhead. Hence, they are primarily applicable to slowly varying channels, where the time overhead in training occupies an insignificant fraction of the channel coherence time. An exception is [13], where, taking the training overhead into account, the authors concluded that for nonzero multiplexing gain g_m , the diversity order saturates as r increases, where r is the number of receive antennas. Hence, for fast varying channels, the authors suggest turning off receive antennas in order to achieve higher multiplexing gains. It is important to account for the training duration overhead in deriving the achievable DMT, because, as the SNR goes to infinity, although the estimation error goes to zero, the training duration overhead remains fixed and has a direct impact on the DMT. Also, by modeling the CSIT as the sum of the true CSI and an additive error, most of the past studies implicitly assume that a channel-agnostic orthogonal training signal is employed for channel estimation.

The authors are with the Dept. of Electrical Communication Eng. at IISc, Bangalore, India. (e-mails: {bharath, cmurthy}@ece.iisc.ernet.in)

The authors gratefully acknowledge the financial support of the DST, India and of the EPSRC, UK under the auspices of the India-UK Advanced Technology Center (IU-ATC), as well as that of the Defense Research Development Organization, India.

This work has appeared in part in [1].

When the training signal is channel-dependent, the imperfect CSI can no longer be modeled as the sum of the true CSI and an additive noise. Due to this, the existing results cannot be directly extended to analyze the DMT performance of channel-dependent training schemes.

When the channel is reciprocal and block-fading, e.g., in a TDD system, the receiver could exploit its channel knowledge (acquired through an initial forward-link training phase) in designing its reverse-training sequence, not only to reduce the channel estimation error at the transmitter, but also to reduce the required training duration overhead. Hence, the goals of this paper are two-fold: (a) to analyze the DMT performance of a *channel dependent* training scheme for acquiring CSIT and an associated power control mechanism for data transmission; and (b) to contrast the DMT performance of the proposed training and power control schemes with that achieved by conventional channel agnostic training schemes. Our study focuses on point-to-point Single Input Multiple Output (SIMO) systems. This is of practical importance, since it applies, for example, to the uplink of wireless networks where the base station has multiple antennas, the mobile users have a single antenna, and orthogonal access is used (e.g., OFDM/TDMA) as in WLANs and 4G/LTE systems. The channel dependent training sequence employed here was first proposed by us in [20] and [1] in a MIMO and SIMO context, respectively, and was independently explored in [21], although not in a DMT context.

In this paper, for analytical simplicity and clarity of presentation, we start by assuming that perfect CSI is available at the receiver, as in [10]–[12]. We propose a fixed-power RCT sequence, using which, the CSI can be estimated at the transmitter using a minimum duration of only one symbol, i.e., with a *factor of r reduction* in training duration compared to orthogonal RCT. For data transmission, we propose a modified truncated channel inversion-type power control scheme based on the noisy CSIT. For this system, we show that a diversity of $d(g_m) = r \left(s + 1 - \frac{g_m L_c}{L_c - L_{B,\tau}} \right)$ is achievable. Here, g_m is the multiplexing gain, L_c is the coherence time, $L_{B,\tau} \geq 1$ is the reverse training duration, and $1 \leq s < r$ is a parameter in the data power control scheme. (See Section III.)

Next, we consider the more practical case where noisy CSIR is acquired via a forward link training sequence, and propose a *three-way* training scheme followed by data transmission. We show that a DMT of $d(g_m) = r(s + 1 - \frac{g_m L_c}{L_c - \beta})$ is achievable, where $\beta \geq 3$ is the total training overhead from all three training phases, which is again an improvement over conventional orthogonal training schemes. For example, a nonzero diversity order can be achieved with $\frac{L_c - (r+2)}{L_c} \leq g_m < \frac{L_c - 3}{L_c}$, which is not possible with orthogonal training schemes without switching off receive antennas and incurring an associated reduction in diversity order. (See Section IV.)

Note that although the perfect CSIR case is a special case of the three-way training scheme with infinite forward-link training power, we briefly present the perfect CSIR case also, as it provides insights into the impact of the reverse-training and data power control mechanisms on the DMT. Moreover, it is useful as an upper bound on the performance

with imperfect CSIR. Also, we assume that power control is employed only at the transmitter and focus on fixed-power RCT in the sequel. Using power controlled RCT significantly changes the problem; we analyze this case in our follow up work [22].

An important implication of our work is that it shows that by exploiting the receiver's knowledge of the CSI in designing the *reverse channel training (RCT)* sequence and using our proposed data power control scheme, one can achieve a higher diversity order than conventional RCT for all values of g_m . Somewhat surprisingly, we also demonstrate that although the DMT analysis corresponds to taking the SNR to infinity, it can nonetheless be used to discriminate between different training schemes both in terms of the estimation error as well as the training overhead. At finite SNR, this translates to an improvement in the outage probability performance and the achievable data rate, as will be illustrated through Monte-Carlo simulations in Section VI.

We use the following notation. Bold face letters are used for vectors and normal font letters are used for scalars. We use $\mathbb{E}(\cdot)$ to denote the expected value of (\cdot) . We use $\|\mathbf{h}\|_2$ to represent the ℓ_2 norm of \mathbf{h} . The transpose conjugate, absolute value, and real part are denoted by $(\cdot)^H$, $|\cdot|$ and $\Re\{\cdot\}$, respectively. We write $f(\bar{P}) \doteq \frac{1}{\bar{P}^k}$ to mean $-\lim_{\bar{P} \rightarrow \infty} \frac{\log f(\bar{P})}{\log \bar{P}} = k$. Similarly, we define $f(\bar{P}) \preceq \frac{1}{\bar{P}^k}$ to mean $-\lim_{\bar{P} \rightarrow \infty} \frac{\log f(\bar{P})}{\log \bar{P}} \geq k$.

II. SYSTEM MODEL

The system model consists of two communicating nodes, *node A* with a single antenna and *node B* with r antennas, with *node A* attempting to send data to *node B* over a wireless channel. The forward channel from *node A* to *node B*, denoted by $\mathbf{h} \in \mathbb{C}^{r \times 1}$, is modeled as a Rayleigh flat fading channel whose entries are *i.i.d.* Circularly Symmetric Complex Gaussian (CSCG) random variables with zero mean and unit variance, i.e., $\mathcal{CN}(0, 1)$. The channel is assumed to be block-fading, i.e., it remains constant for a duration of the coherence time L_c , and evolve in an *i.i.d.* fashion across coherence times. We assume a TDD system with perfect reciprocity, and hence, taking the complex conjugate of the received signal at *node A*, the reverse link channel is \mathbf{h}^H . We let $\mathbf{h} = \sigma \mathbf{v}$, where $\sigma = \|\mathbf{h}\|_2$ is the singular value and $\mathbf{v} \triangleq \frac{\mathbf{h}}{\|\mathbf{h}\|_2}$ is the singular vector of \mathbf{h} . Since our goal is to study the achievable DMT performance with channel training, we first explain the two-way training protocol used for acquiring CSI at *node B* and *node A*. Later, in Sec. IV, an additional phase of forward link training is introduced, which is not presented here for simplicity.

1) Phase I (Forward-link training): Here, the training sequence $x_{A,\tau} = \sqrt{P} L_{A,\tau_1}$ is transmitted from *node A* to *node B*, where L_{A,τ_1} denotes the training duration and \bar{P} is the training power¹. Throughout this paper, we use \bar{P} as the average power constraint during both training and data

¹Strictly speaking, $x_{A,\tau} = \sqrt{\bar{P}}$ is transmitted repeatedly L_{A,τ_1} times. Mathematically, this is equivalent to using $x_{A,\tau} = \sqrt{\bar{P} L_{A,\tau_1}}$ for a duration of one unit.

transmission. The corresponding received training signal is given by,

$$\mathbf{y}_{B,\tau} = \mathbf{h}\sqrt{\bar{P}L_{A,\tau_1}} + \mathbf{w}_{B,\tau}. \quad (1)$$

The entries of $\mathbf{w}_{B,\tau} \in \mathcal{C}^{r \times 1}$ are assumed to be distributed as *i.i.d.* $\mathcal{CN}(0,1)$. From the received training signal $\mathbf{y}_{B,\tau}$, *node B* computes an MMSE estimate of \mathbf{h} , denoted $\hat{\mathbf{h}}$. The error in the estimate, denoted $\tilde{\mathbf{h}} \triangleq \mathbf{h} - \hat{\mathbf{h}}$, has *i.i.d.* $\mathcal{CN}(0, 1/(1 + \bar{P}L_{A,\tau_1}))$ distributed entries.

In a TDD-SIMO system, *node A* only requires knowledge of σ to perform power control, which in turn improves the diversity order compared to the no-CSIT case. Therefore, in phase II, we estimate only σ at *node A*, using a channel dependent training sequence.

2) **Phase II (Reverse-link training)**: Since *node B* has an estimate (say, $\hat{\mathbf{v}} \triangleq \frac{\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2}$) of the channel, in this phase, it exploits its CSI to transmit the following training sequence [1], [20]:

$$\mathbf{x}_{B,\tau} = \sqrt{\bar{P}L_{B,\tau}}\hat{\mathbf{v}}, \quad (2)$$

where $L_{B,\tau}$ is the reverse training duration. Using the corresponding received signal, $y_{A,\tau} \triangleq \mathbf{h}^H \mathbf{x}_{B,\tau} + w_{A,\tau}$, where the AWGN $w_{A,\tau} \in \mathcal{C}$ is distributed as $\mathcal{CN}(0,1)$, *node A* computes an estimate of the singular value as follows:

$$\hat{\sigma} \triangleq \frac{\Re\{y_{A,\tau}\}}{\sqrt{\bar{P}L_{B,\tau}}} = \sigma \Re\{\mathbf{v}^H \hat{\mathbf{v}}\} + \bar{w}_{A,\tau}, \quad (3)$$

where $\bar{w}_{A,\tau} \triangleq \frac{\Re\{w_{A,\tau}\}}{\sqrt{\bar{P}L_{B,\tau}}}$. Note that the estimate $\hat{\sigma}$ could be negative; this is taken care of by the power control proposed in Sec. III, which uses $\hat{\sigma}$ only when it is greater than a positive threshold. Since a low or negative $\hat{\sigma}$ is likely to be inaccurate, the thresholding technique helps to avoid the poor DMT performance due to such estimates. The RCT scheme employed above is different from existing channel agnostic methods in that the minimum training length in the proposed scheme is only 1 symbol. This represents a factor of r reduction compared to orthogonal RCT schemes, where the minimum training length increases linearly with r , and this difference in overhead could be significant when L_c is small. Also, if $\hat{\mathbf{v}}$ is error-free, it is the optimal beamforming vector for estimating σ at *node A*.

3) **Multiplexing Gain and Diversity Order**: We recall the definitions of the multiplexing gain, g_m , and the diversity order d from [2]:

$$g_m \triangleq \lim_{\bar{P} \rightarrow \infty} \frac{R_{\bar{P}}}{\log \bar{P}}, \quad d \triangleq - \lim_{\bar{P} \rightarrow \infty} \frac{\log P_{out}}{\log \bar{P}}, \quad (4)$$

where $R_{\bar{P}}$ is the target data rate when the average data power constraint is \bar{P} , and P_{out} is the corresponding outage probability, i.e., the probability that $R_{\bar{P}}$ exceeds the channel capacity. In this work, the target data rate $R_{\bar{P}} = g_m \log \bar{P}$ is fixed and is independent of the CSIT; the extension of our proposed methods to joint rate and power adaptation is relegated to future work. The rate of data transmission $R_{\bar{P}}$ is increased with \bar{P} by increasing the cardinality of the signal set, keeping the symbol duration fixed. We ignore the effect of spectral leakage, and assume that the signal bandwidth remains fixed as \bar{P} goes to infinity. Also, we use outage probability as

a proxy for the probability of error at high SNR with finite-length codes; this is because the probability of error can be made to decrease as fast as the outage probability using finite-length approximately universal codes [23], [24].

In the next section, we assume perfect CSI at *node B* and derive the achievable DMT performance of our proposed training and data transmission schemes.

III. DMT ANALYSIS WITH PERFECT CSIR

When the CSIR is perfect, we have $\hat{\mathbf{v}} = \mathbf{v}$, and in this case, it is easy to see that (2) is optimal for estimating σ given a power constraint \bar{P} on the training signal. This is because, in general, the training signal can be expressed as the linear combination $\mathbf{x}_{B,\tau} = \delta \mathbf{v} + \beta \mathbf{v}_\perp$, where \mathbf{v}_\perp is orthogonal to \mathbf{v} and δ and β are some constants. Then, the received training signal at *node A* is $y_{A,\tau} = \delta \sigma + w_{A,\tau}$, i.e., the power in \mathbf{v}_\perp does not help in estimating σ . From (3), an *unbiased* estimator of the singular value at *node A* is given by

$$\hat{\sigma} = \sigma + \bar{w}_{A,\tau}. \quad (5)$$

Note that since the channel is assumed to be Rayleigh fading, σ^2 is chi-square distributed with $2r$ degrees of freedom. Also, we employ this estimator primarily because we are interested in deriving the achievable DMT performance, and for this purpose, this simple unbiased estimator is sufficient.

A. Power-Controlled Data Transmission from Node A to Node B

Given the CSIT $\hat{\sigma}$ in (5), *node A* uses a power $\mathcal{P}(\hat{\sigma})$ in the forward link *data transmission* phase, to avoid outages while satisfying the average data power constraint \bar{P} . The corresponding *data signal* received at *node B* is given by,

$$\mathbf{y}_{B,d} = \sqrt{\mathcal{P}(\hat{\sigma})}\mathbf{h}x_{A,d} + \mathbf{w}_{B,d}, \quad (6)$$

where $x_{A,d} \sim \mathcal{CN}(0,1)$, and with appropriate power normalization, the entries of the AWGN $\mathbf{w}_{B,d} \in \mathcal{C}^{r \times 1}$ are assumed to be *i.i.d.* $\mathcal{CN}(0,1)$. Also, $\mathcal{P}(\hat{\sigma})$ is chosen independent of $x_{A,d}$ such that $\mathbb{E}\{\mathcal{P}(\hat{\sigma})\} = \bar{P}$, where the expectation is with respect to $\hat{\sigma}$ given in (5), taken across all coherence blocks. Since $\mathbb{E}\{|x_{A,d}|^2\} = 1$ within a block, this ensures that the average data power constraint at *node A* is satisfied.

We now present the data power control function $\mathcal{P}(\hat{\sigma})$ considered in this paper. Our proposed power control function is motivated as follows. The capacity of a fading channel with *mismatched* CSIT and CSIR is not known in closed form [25]. Since the outage probability computation requires a closed form expression for the capacity, we consider a genie-aided receiver as in [26], where *node B* is assumed to know $\mathcal{P}(\hat{\sigma})$. This is schematically illustrated in Fig. 1. Then, the achievable data rate conditioned on the knowledge of $\sqrt{\mathcal{P}(\hat{\sigma})}\mathbf{h}$ is given by [25]

$$C \triangleq \frac{L_c - L_{B,\tau}}{L_c} \log(1 + \sigma^2 \mathcal{P}(\hat{\sigma})). \quad (7)$$

An outage occurs when $R_{\bar{P}}$, the target data rate, exceeds C . Its probability is upper bounded by

$$P_{out} \triangleq \Pr\left(\frac{L_c - L_{B,\tau}}{L_c} \log(1 + \sigma^2 \mathcal{P}(\hat{\sigma})) < R_{\bar{P}}\right). \quad (8)$$

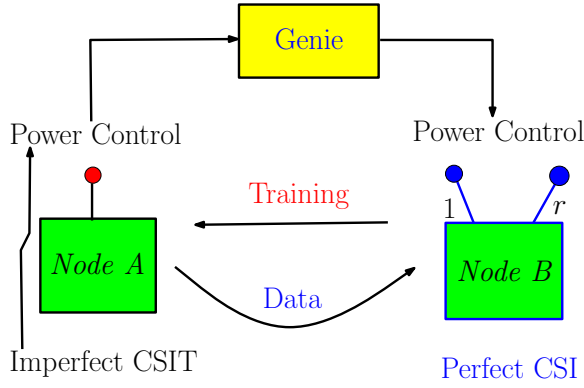


Fig. 1. System model for reverse channel training with perfect CSIR used in Section III.

Note that the exact outage probability is obtained by minimizing the right hand side above over all $\mathcal{P}(\hat{\sigma})$ satisfying $\mathbb{E}\{\mathcal{P}(\hat{\sigma})\} = \bar{P}$. Hence, using our proposed data power control scheme leads to an upper bound on the outage probability, which is sufficient for obtaining the achievable DMT performance. If the CSIT is perfect (i.e., $\hat{\sigma}^2 = \sigma^2$), it is shown in [3] that the power control that minimizes the outage probability is given by

$$\Phi(\sigma^2) \triangleq \frac{\exp\left(\frac{R_{\bar{P}} L_c}{L_c - L_{B,\tau}}\right) - 1}{\sigma^2}. \quad (9)$$

Note that since $R_{\bar{P}} = g_m \log \bar{P}$ and $\mathbb{E}\{\frac{1}{\sigma^2}\} = \frac{1}{r-1}$, $\Phi(\sigma^2)$ satisfies $\mathbb{E}\{\Phi(\sigma^2)\} \leq \bar{P}$ for large enough \bar{P} , provided $g_m \leq (L_c - L_{B,\tau})/L_c$. With inaccurate CSIT, due to the estimation error in $\hat{\sigma}$, the natural extension of using a transmission power of $\Phi(\hat{\sigma}^2)$ could result in allocating insufficient power or more power than required, which could lead to suboptimal performance. Also, inverting the channel for all values of $\hat{\sigma}$ results in an infinite average power since the Gaussian noise can make the estimate $\hat{\sigma}$ arbitrarily small with a non-zero probability. One solution is to use a transmit power of $\Phi(\hat{\sigma}^2)$ when $\hat{\sigma} > \theta_0$ and a zero power otherwise, where θ_0 is chosen such that $\mathbb{E}[\Phi(\hat{\sigma}^2)1_{\hat{\sigma} > \theta_0}] = \bar{P}$. The drawback of this method is that it results in an outage probability of 1 when $\hat{\sigma} \leq \theta_0$, leading to a zero diversity order. To overcome this problem, we choose the threshold θ_0 such that $\theta_0 \rightarrow 0$ as $\bar{P} \rightarrow \infty$. Moreover, when $\hat{\sigma} \leq \theta_0$, we do not necessarily want to use zero power, since the small value of $\hat{\sigma}$ could be due to the estimation error. This motivates the following modified power control:

$$\mathcal{P}(\hat{\sigma}) \triangleq \begin{cases} \bar{P}^l & \hat{\sigma} \leq \theta_{\bar{P}}, \\ \kappa_{\bar{P}} \times \Phi(\hat{\sigma}^{2s}) & \hat{\sigma} > \theta_{\bar{P}}, \end{cases} \quad (10)$$

where $s \geq 1$ is a parameter, and we use $\theta_{\bar{P}} \triangleq \frac{1}{\bar{P}^n}$, $n > 0$, for mathematical tractability. The parameters n , $\kappa_{\bar{P}}$ and $l > 0$ are chosen such that $\mathbb{E}[\mathcal{P}(\hat{\sigma})] = \bar{P}$. Although similar power control schemes have been employed in the literature with perfect CSIT [3] or orthogonal RCT [12], [13], [19], the form in (10) is new. Specifically, the power control scheme in [3], [12], [13] can be obtained from (10) by setting $s = 1$, $\theta_{\bar{P}} = 0$ and $l = -\infty$; while that in [19] can be obtained by setting $s = r$, $\theta_{\bar{P}} = 0$ and $l = -\infty$.

Power constraint: The description of the power control would be complete if the parameters n , $\kappa_{\bar{P}}$ and l can be chosen such that $\mathbb{E}[\mathcal{P}(\hat{\sigma})] = \bar{P}$, which is the essence of the following Lemma.

Lemma 1: Let $\theta_{\bar{P}} \triangleq \frac{1}{\sqrt{\bar{P}}}$. For $1 \leq s < r$, there exists a $\kappa_{\bar{P}} \triangleq \frac{1}{\bar{P}^{\frac{1}{\alpha}} - 1}$, where $\alpha \triangleq \frac{L_c - L_{B,\tau}}{L_c}$, such that $\mathbb{E}[\mathcal{P}(\hat{\sigma})] = \bar{P}$, if $0 \leq l \leq r + 1$.

Proof: See Appendix B. ■

Due to Lemma 1, in the rest of this paper, we consider $\theta_{\bar{P}} = 1/\sqrt{\bar{P}}$. Also, in Sec. IV, we show that a minor modification of the above data power control scheme can be employed even with imperfect CSIR. The next subsection presents the achievable DMT of the proposed training and power control schemes.

B. Achievable DMT Analysis

Theorem 1: Given r receive antennas and $L_{B,\tau}$ training symbols being used per coherence interval L_c to estimate the CSIT in a SIMO system with perfect CSIR and a genie-aided receiver, an achievable diversity order as a function of multiplexing gain g_m is given by

$$d(g_m) = r \left(\min\{l, s + 1\} - \frac{g_m}{\alpha} \right), \quad (11)$$

where $0 \leq l \leq r + 1$, $1 \leq s < r$, $0 \leq g_m < \alpha$, and $\alpha \triangleq \frac{L_c - L_{B,\tau}}{L_c}$ represents the fractional data transmit duration.

Proof: See Appendix C. ■

Remark: From a DMT perspective, it is clear from Theorem 1 that $s \rightarrow r, l = r + 1$ is superior to $s = 1, l = 2$. On the other hand, when $\hat{\sigma} < 1$, $\Phi(\hat{\sigma}^{2r})$ could be much greater than $\Phi(\hat{\sigma}^2)$. Thus, in practical systems with a peak power per transmitted codeword constraint, $s = 1, l = 2$ could be preferable over $s \rightarrow r, l = r + 1$. In the sequel, for convenience, we associate $l = 2$ with $s = 1$ and $l = r + 1$ with $s \rightarrow r$, and drop the explicit dependence of the diversity order on l . Further remarks and discussions on the result obtained here are deferred to Sec. V.

IV. THREE WAY TRAINING

In this section, we consider the more practical scenario where training is performed in both directions. We show that with fixed power training, one can achieve nearly the same DMT as derived in Sec. III for the perfect CSIR case. Unlike in the previous section, the analysis presented here is exact, in the sense that it does not require the assumption of a genie aided receiver, and hence, the DMT derived here is indeed achievable in practice. The transmission protocol now consists of four phases, as shown in Table I. The CSIR and CSIT are obtained by transmitting a fixed power training sequence in both directions, as explained in Sec. II. However, even a small mismatch in the CSI knowledge at *node A* and *node B* can potentially lead to a large mismatch in their estimate of the data transmit power [13]. Thus, it is essential to train *node B* about *node A*'s knowledge of $\mathcal{P}(\hat{\sigma})$. This leads to a third phase of training, which is an additional power-controlled forward-link training phase. First, in the following subsection, we explain the power control scheme that is employed here.

A. Power Control Scheme

The power control scheme we propose to employ in this section is as given by (10), due to the following. Let $\hat{\mathbf{h}}$ denote the MMSE estimate of the channel at *node B*, and consider $\hat{\sigma}$ in (3). We have

$$\hat{\sigma} \triangleq \frac{\Re\{y_{A,\tau}\}}{\sqrt{PL_{B,\tau}}} = \Re\{\hat{\mathbf{h}}^H \hat{\mathbf{v}}\} + \Re\{\tilde{\mathbf{h}}^H \hat{\mathbf{v}}\} + \frac{\Re\{w_{A,\tau}\}}{\sqrt{PL_{B,\tau}}} \\ = \|\hat{\mathbf{h}}\|_2 + \tilde{w}_{eff}, \quad (12)$$

where $\tilde{w}_{eff} \triangleq \Re\{\tilde{\mathbf{h}}^H \hat{\mathbf{v}}\} + \frac{\Re\{w_{A,\tau}\}}{\sqrt{PL_{B,\tau}}}$. Note that $\hat{\mathbf{h}}$ and $\tilde{\mathbf{h}}$ are independent Gaussian random variables². Since $\hat{\mathbf{v}}$ is uniformly distributed on the unit sphere and is independent of $\hat{\mathbf{h}}$, $\Re\{\tilde{\mathbf{h}}^H \hat{\mathbf{v}}\}$ is Gaussian distributed. This implies that the effective noise, \tilde{w}_{eff} , is Gaussian distributed with $\mathbb{E}|\tilde{w}_{eff}|^2 \doteq \frac{1}{P}$ and independent of $\hat{\mathbf{h}}$. Therefore, the estimate of the singular value at *node A* is statistically similar to the estimate given by (5) in the perfect CSIR case. Thus, we use a similar power control, $\mathcal{P}(\hat{\sigma})$ in (10), where $\hat{\sigma}$ is given by (12). Also, with a slight abuse of notation, $\alpha \triangleq \frac{L_c - L_{B,\tau} - L_{A,\tau_1} - L_{A,\tau_2}}{L_c}$, where L_{A,τ_2} is the training duration in the third phase of training (phase III), which is in the forward-link direction.

In this section, without loss of generality, we move the power scaling \sqrt{P} into the data symbol transmitted by *node A*, so that $\mathbb{E}\{\mathcal{P}(\hat{\sigma})\} = 1$ (see (13) below), where the expectation is taken with respect to the distribution of $\hat{\sigma}$ in (12). Now, in the proof of Lemma 1, using the probability density function (pdf) of $\|\hat{\mathbf{h}}\|_2$ in place of the pdf of σ , and noting that the effective noise variance $\doteq 1/P$, we get $\kappa_{\bar{P}} \doteq \frac{1}{P_{gm}/\alpha}$ and the constraint $0 \leq l \leq r$ to satisfy $\mathbb{E}\{\mathcal{P}(\hat{\sigma})\} = 1$ at high SNR. In the next subsection, we explain the third round of training that alleviates the mismatch in the knowledge of the data transmit power.

B. Phase III (Power-Controlled Forward Link Training)

In this phase, *node A* transmits the training sequence: $x_{A,\tau_2} = \sqrt{PL_{A,\tau_2}}\sqrt{\mathcal{P}(\hat{\sigma})}$, where L_{A,τ_2} is the training duration. The corresponding received training signal at *node B* is given by,

$$\mathbf{y}_{B,\tau_2} = \sqrt{\bar{P}L_{A,\tau_2}}\sqrt{\mathcal{P}(\hat{\sigma})}\mathbf{h} + \mathbf{w}_{B,\tau_2}, \quad (13)$$

where $\mathbf{w}_{B,\tau_2} \in \mathbb{C}^{r \times 1}$ is the AWGN with $\mathcal{CN}(0, 1)$ entries. The goal at *node B* is to estimate the composite channel $\mathbf{p}_c \triangleq \sqrt{\mathcal{P}(\hat{\sigma})}\mathbf{h}$. Dividing (13) by $\sqrt{PL_{A,\tau_2}}$, we get

$$\tilde{\mathbf{y}}_{B,\tau_2} \triangleq \frac{\mathbf{y}_{B,\tau_2}}{\sqrt{PL_{A,\tau_2}}} = \mathbf{p}_c + \frac{\mathbf{w}_{B,\tau_2}}{\sqrt{PL_{A,\tau_2}}}. \quad (14)$$

² $\hat{\mathbf{h}} \rightarrow \mathbf{h}$ as $\bar{P} \rightarrow \infty$. Moreover, $\|\hat{\mathbf{h}}\|_2$ is a chi distributed random variable.

From (14), *node B* computes an MMSE estimate of \mathbf{p}_c , denoted by $\hat{\mathbf{p}}_c$. Let $\tilde{\mathbf{p}}_c \triangleq \mathbf{p}_c - \hat{\mathbf{p}}_c$. Although a closed form expression for $\hat{\mathbf{p}}_c$ is hard to find, the error $\tilde{\mathbf{p}}_c$ in the MMSE estimate has the following interesting property, which facilitates the calculation of the outage probability in Sec. IV-D. An analogous result has been shown in [27] for the scalar case.

Lemma 2: $\mathbb{E}\|\tilde{\mathbf{p}}_c\|_2^{2z} \preceq \frac{1}{P^z}$ for every $z > 0$.

Proof: See Appendix D. ■

C. Phase IV (Data Transmission)

Using $\mathcal{P}(\hat{\sigma})$, *node A* sends the data signal $x = \sqrt{P\mathcal{P}(\hat{\sigma})}x_{A,d}$, where $x_{A,d}$ is distributed as $\mathcal{CN}(0, 1)$ and is independent of $\mathcal{P}(\hat{\sigma})$. Note that $\mathbb{E}|x|^2 = \bar{P}$ by construction, where the expectation is taken with respect to both $\hat{\sigma}$ and $x_{A,d}$. The corresponding signal received at *node B* is

$$\mathbf{y}_{B,d} = \sqrt{\bar{P}\mathcal{P}(\hat{\sigma})}\mathbf{h}x_{A,d} + \mathbf{w}_{B,d} \quad (15)$$

$$= \sqrt{\bar{P}}\hat{\mathbf{p}}_cx_{A,d} + \sqrt{\bar{P}}\tilde{\mathbf{p}}_cx_{A,d} + \mathbf{w}_{B,d}. \quad (16)$$

Since $\hat{\mathbf{p}}_c$ is an MMSE estimate, using the worst case noise theorem [5], we have the following lower bound on the mutual information, $I(x_{A,d}; \mathbf{y}_{B,d} | \hat{\mathbf{p}}_c) \geq C_{AB}$, where

$$C_{AB} \triangleq \alpha \log \left(1 + \frac{\bar{P}\|\hat{\mathbf{p}}_c\|_2^2}{\frac{\bar{P}}{r}\mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2|\tilde{\mathbf{y}}_{B,\tau_2}] + 1} \right), \quad (17)$$

and $\alpha \triangleq \frac{L_c - L_{B,\tau} - L_{A,\tau_1} - L_{A,\tau_2}}{L_c}$ is the fractional data transmit duration after accounting for the time overheads in all three training phases.

D. DMT Analysis With Three-Way Training

Theorem 2: For a SIMO system with r receive antennas and three phases of training and the data transmission phase as described in Table I, an achievable DMT is given by

$$d(g_m) = r \left(\min\{l, s\} + 1 - \frac{g_m}{\alpha} \right), \quad (18)$$

where $0 \leq l \leq r$, $1 \leq s < r$, $0 \leq g_m < \alpha$, and $\alpha \triangleq \frac{L_c - L_{B,\tau} - L_{A,\tau_1} - L_{A,\tau_2}}{L_c}$.

Proof: See Appendix E. ■

Remark: The above three way training scheme can be generalized to k training rounds to improve the diversity order, as in [17], [19]. However, this is mathematically cumbersome and out of the scope of our work.

TABLE I
THREE WAY TRAINING IN A TDD-SIMO SYSTEM

Phase	Description	Input-Output Equation
I	Fixed power training (<i>Node A</i> \rightarrow <i>Node B</i>)	$\mathbf{y}_{B,\tau} = \mathbf{h}x_{A,\tau} + \mathbf{w}_{B,\tau}$
II	Fixed power training (<i>Node B</i> \rightarrow <i>Node A</i>)	$y_{A,\tau} = \mathbf{h}^H \mathbf{x}_{B,\tau} + w_{A,\tau}$
III	Power controlled training (<i>Node A</i> \rightarrow <i>Node B</i>)	$\mathbf{y}_{B,\tau_2} = \sqrt{\bar{P}L_{A,\tau_2}}\mathcal{P}(\hat{\sigma})\mathbf{h} + \mathbf{w}_{B,\tau_2}$
IV	Power controlled data (<i>Node A</i> \rightarrow <i>Node B</i>)	$\mathbf{y}_{B,d} = \mathbf{h}x_{A,d} + \mathbf{w}_{B,d}$

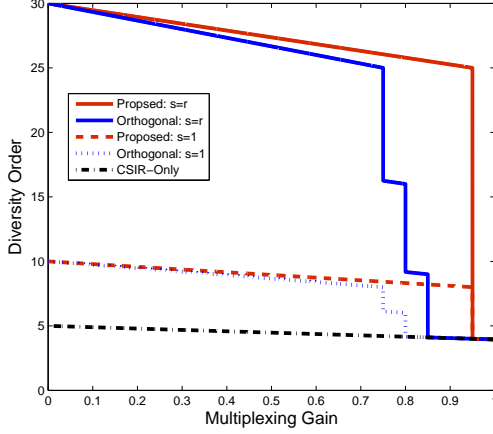


Fig. 2. The achievable DMT with the training and power control scheme proposed in Sec. III, compared with the performance with orthogonal RCT and the data power control proposed in [13], [19] (and appropriately accounting for the training duration overhead and switching off antennas to achieve higher values of g_m). The plot corresponds to a SIMO system with $r = 5$ antennas, with coherence time $L_c = 20$ symbols, and reverse training duration of $L_{B,\tau} = 1$ symbol.

V. DISCUSSION

Recall that with perfect CSIR and imperfect CSIT, with $l \geq s + 1$, and for a genie aided channel, it was shown in Theorem 1 that the following DMT is achievable:

$$d(g_m) = r \left[s + 1 - \left(\frac{g_m L_c}{L_c - L_{B,\tau}} \right) \right], \quad (19)$$

where $1 \leq s < r$, $0 \leq g_m \leq \frac{L_c - L_{B,\tau}}{L_c}$. In contrast, for the same genie aided channel, it was shown in [26] that a diversity order of

$$d_s(g_m) = r \left[2 - \left(\frac{g_m L_c}{L_c - r L_{B,\tau}} \right) \right], \quad 0 \leq g_m \leq \frac{L_c - r L_{B,\tau}}{L_c} \quad (20)$$

is achievable using orthogonal reverse channel training. Note that $d_s(g_m)$ saturates as r gets large, as opposed to (19), which is monotonically increasing in r . In order to achieve a $g_m > \frac{L_c - r L_{B,\tau}}{L_c}$, in [13], the authors suggest turning off one receive antenna at a time to reduce the training burden until $r = 2$. For example, turning off one antenna, $g_m \in \left[\frac{L_c - r L_{B,\tau}}{L_c}, \frac{L_c - (r-1)L_{B,\tau}}{L_c} \right]$ is achievable at a reduced diversity order of $d_s(g_m) = (r-1) \left[2 - \left(\frac{g_m L_c}{L_c - (r-1)L_{B,\tau}} \right) \right]$. This is in contrast to our result, which can accommodate a larger multiplexing gain, $g_m \leq \frac{L_c - L_{B,\tau}}{L_c}$ irrespective of r , while simultaneously achieving a higher diversity order at each g_m . We note that for a SIMO channel, a diversity order of $r(r+1-g_m)$ for $0 \leq g_m < 1$ was obtained in [12], [19], using channel-independent training, and without accounting for the training duration overhead. This corresponds to taking $L_c \rightarrow \infty$ in (19). The performance of the proposed scheme is schematically contrasted with orthogonal RCT in Fig. 2 for a SIMO system with $r = 5$, $L_c = 20$, and $L_{B,\tau} = 1$ symbol. The advantage of the proposed scheme at higher values of the multiplexing gain is clear from the plot. The

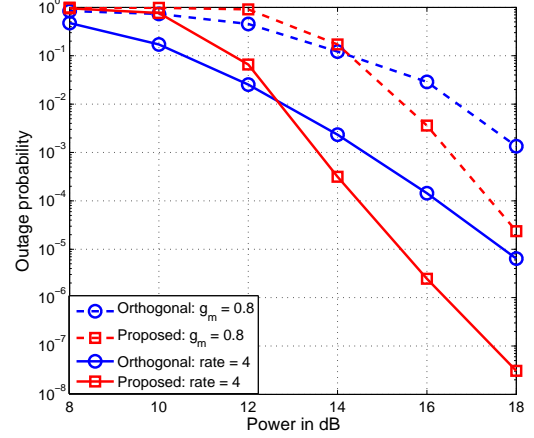


Fig. 3. Outage probability versus the average data power \bar{P} for the fixed-power training scheme proposed in Sec. III, with the data power control scheme given by (10) with $s = 1$. Here, $r = 3$, $L_c = 40$ and $L_{B,\tau} = 1$. With $g_m = 0.8$, the target data rate was set as $R_{\bar{P}} = 4 + g_m \log \bar{P}$ to facilitate the comparison of the curves.

proposed training scheme thus results in a *factor r -reduction* in the training duration, which, along with the proposed data power control scheme, translates to an increase in the range of achievable multiplexing gains, while simultaneously offering a better diversity order compared to orthogonal RCT schemes.

Comparing Theorems 1 and 2, we see that the DMT performance of a genie aided receiver with perfect CSIR is an upper bound on the performance of the system with imperfect CSIR and CSIT, as expected. Also, the performance of the two systems is similar, except that in the latter case, the factor α captures the loss in data transmission time due to all three training phases. Similar observations as the above regarding the improvement in DMT can be made for the three way training scheme compared to orthogonal RCT schemes.

VI. SIMULATION RESULTS

We now briefly present Monte-Carlo simulation results to illustrate the outage probability performance of our proposed RCT and forward-link data power control schemes. We consider a Rayleigh fading channel with three receive antennas. We calculate the outage probability by averaging over 10^8 *i.i.d.* channel and training noise instantiations. We set the channel coherence time and reverse training duration as $L_c = 40$ and $L_{B,\tau} = 1$, respectively. Figures 3 and 4 show the outage probability of the proposed fixed-power training scheme and the data power control scheme in (10) with $s = 1$ and $s = r = 3$, respectively, as a function of \bar{P} , with $g_m = 0$ and $R_{\bar{P}} = 4$ bits/channel use (1 and 1.5 bits/channel use in case of Fig. 4), and with $g_m = 0.8$. Although the slopes of the curves do not match with the theoretical diversity order because the latter requires infinite SNR, the improved performance of the proposed schemes is clear from the graphs. Also, in Fig. 3, since the proposed scheme uses only $L_{B,\tau} = 1$ training symbol while the orthogonal RCT scheme uses $r L_{B,\tau} = 3$ training symbols, the former shows a higher outage than the latter at lower SNRs. Note that, we have not plotted the outage

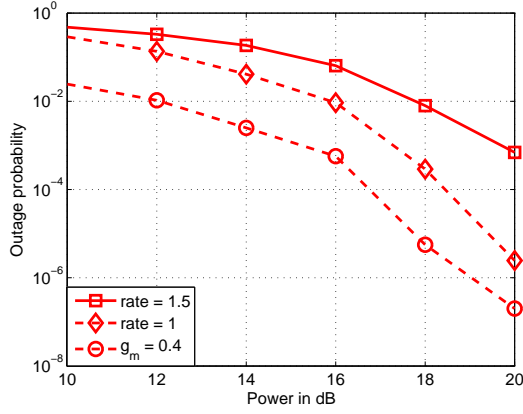


Fig. 4. Outage probability versus the average data power \bar{P} for the fixed-power training scheme proposed in Sec. III, with the data power control scheme given by (10) with $s = r$. Here, $r = 3$, $L_c = 40$ and $L_{B,\tau} = 1$.

performance of the three-way training scheme in Sec. IV. This is because the outage probability is hard to compute, since a closed-form expression for $\hat{\mathbf{p}}_c$ is not available.

VII. CONCLUSIONS

This paper proposed reverse training and data power control schemes for a TDD-SIMO system with perfect/imperfect CSIR and investigated its DMT performance. It was shown that a diversity order of $d(g_m) = r(s + 1 - \frac{g_m}{\alpha})$ is achievable for $l \geq s + 1$, $1 \leq s < r$ and $0 \leq g_m < \alpha$, where α represents the fractional data transmit duration. In contrast to channel agnostic orthogonal training schemes, the diversity order was shown to increase monotonically with r at nonzero multiplexing gain, which is a significant improvement. The DMT analysis was extended to a more practical situation where the training is done in both directions. In this case also, it was shown that the DMT performance can improve quadratically with the number of receive antennas, and nearly the same DMT can be achieved as that with perfect CSIR and a genie-aided receiver. In terms of system design for reciprocal SIMO systems, the key messages from this work are that it is important to (a) exploit the CSI at the receiver in designing the RCT and (b) use a modified channel-inversion type power control scheme that transmits data at some non-zero power even when the estimated singular value at the transmitter is poor. For fast varying channels, these ingredients can lead to a significant advantage in DMT performance, which, at finite SNR, can translate to a large improvement in outage probability performance compared to orthogonal training schemes. Future work could extend the DMT analysis to a time-selective block fading reciprocal channel, where the channel is correlated within a block [28].

APPENDIX

A. Useful Lemmas

Lemma 3: If the random variable σ^2 is a chi-square distributed with $2r$ degrees of freedom, then $\Pr\{\sigma^2 < z\} \leq \frac{z^r}{r!}$, $z \geq 0$.

Proof: The result follows from

$$\Pr\{\sigma^2 < z\} = \frac{1}{(r-1)!} \int_0^z e^{-x} x^{r-1} dx \quad (21)$$

$$\leq \frac{1}{(r-1)!} \int_0^z x^{r-1} dx = \frac{z^r}{r!}. \quad (22)$$

Lemma 4: For the system in (3), $|\hat{\sigma}| \leq \hat{\sigma}_U$, where $\hat{\sigma}_U^2 \triangleq (\sigma + |\bar{w}_{A,\tau}|)^2$, with $\bar{w}_{A,\tau} \triangleq \frac{\Re\{w_{A,\tau}\}}{\sqrt{\bar{P}L_{B,\tau}}}$.

Proof: We upper bound the absolute value of (3) as follows:

$$|\hat{\sigma}| \stackrel{(a)}{\leq} \sigma |\Re\{\mathbf{v}^H \hat{\mathbf{v}}\}| + \left| \frac{\Re\{w_{A,\tau}\}}{\sqrt{\bar{P}L_{B,\tau}}} \right| \stackrel{(b)}{\leq} \sigma + |\bar{w}_{A,\tau}|, \quad (23)$$

where (a) follows from the triangle inequality and (b) follows since $|\Re\{\mathbf{v}^H \hat{\mathbf{v}}\}| \leq 1$. ■

B. Proof of Lemma 1

Consider the following constraint on the data power

$$\mathbb{E}[\mathcal{P}(\hat{\sigma})] = \int_{-\infty}^{\infty} \mathcal{P}(\hat{\sigma}) f_{\hat{\sigma}}(\hat{\sigma}; \bar{P}) d\hat{\sigma} = \bar{P}, \quad (24)$$

where $f_{\hat{\sigma}}(\hat{\sigma}; \bar{P})$ is the pdf of $\hat{\sigma}$. Substituting (10) in (24), we get

$$\mathbb{E}[\mathcal{P}(\hat{\sigma})] = \kappa_{\bar{P}} \left[\exp\left(\frac{L_c R_{\bar{P}}}{L_c - L_{B,\tau}}\right) - 1 \right] F(\bar{P}) + I_{\bar{P}}, \quad (25)$$

where $R_{\bar{P}}$ is the target data rate and the data transmit power is \bar{P} ,

$$F(\bar{P}) \triangleq \int_{\theta_{\bar{P}}}^{\infty} \frac{1}{x^{2s}} f_{\hat{\sigma}}(x; \bar{P}) dx \quad \text{and} \quad I_{\bar{P}} \triangleq \bar{P}^l \int_{-\infty}^{\theta_{\bar{P}}} f_{\hat{\sigma}}(x; \bar{P}) dx. \quad (26)$$

The proof is complete by choosing

$$\kappa_{\bar{P}} = \frac{1}{\left(\exp\left(\frac{L_c R_{\bar{P}}}{L_c - L_{B,\tau}}\right) - 1\right) F(\bar{P})} (\bar{P} - I_{\bar{P}}), \quad (27)$$

and showing that $I_{\bar{P}} < \bar{P}$ and that $F(\bar{P})$ is bounded for large \bar{P} when $0 \leq l \leq r + 1$ and $n = 1/2$. From (26), $I_{\bar{P}} = \bar{P}^l \Pr\{\sigma + \bar{w}_{A,\tau} < \theta_{\bar{P}}\}$ can be bounded as,

$$\begin{aligned} I_{\bar{P}} &\stackrel{(a)}{\leq} \frac{\bar{P}^l}{r!} \mathbb{E}(\theta_{\bar{P}} - \bar{w}_{A,\tau})^{2r} \stackrel{(b)}{=} \frac{\bar{P}^l}{r!} \mathbb{E} \sum_{j=0}^r \theta_{\bar{P}}^{2(r-j)} \binom{2r}{2j} \bar{w}_{A,\tau}^{2j} \\ &\stackrel{(c)}{=} \bar{P}^l \max_{j \in \{0,1,\dots,r\}} \frac{1}{\bar{P}^{2(r-j)n+j}} \stackrel{(d)}{=} \frac{1}{\bar{P}^{r-l}}, \end{aligned} \quad (28)$$

where (a) follows from Lemma 3 above, and the expectation is with respect to the distribution of $\bar{w}_{A,\tau}$, (b) follows from the binomial expansion and the fact that $\mathbb{E}\bar{w}_{A,\tau}^i = 0$ when i is odd, (c) follows from $\theta_{\bar{P}} \doteq \frac{1}{\bar{P}^n}$ and $\mathbb{E}\bar{w}_{A,\tau}^{2j} \doteq \frac{1}{\bar{P}^j}$, and (d) follows by substituting $n = 1/2$ in the left hand side. From (28), clearly, $I_{\bar{P}} < \bar{P}$ for large \bar{P} if $l < r + 1$ and $n = 1/2$. When $l = r + 1$ and $n = 1/2$, we have $I_{\bar{P}} \leq \bar{P}$, and therefore we can ensure that $I_{\bar{P}} < \bar{P}$ for large \bar{P} by scaling $I_{\bar{P}}$ by an appropriately chosen constant scaling factor.

Next, we show that $F(\bar{P})$ is bounded. Note that

$$F(\bar{P}) = \int_{\theta_{\bar{P}}}^1 \frac{1}{x^{2s}} f_{\hat{\sigma}}(x; \bar{P}) dx + \int_1^{\infty} \frac{1}{x^{2s}} f_{\hat{\sigma}}(x; \bar{P}) dx. \quad (29)$$

Now, it is sufficient to show that the first integral in (29) is bounded, since the second integral is clearly < 1 . To this end, we need the distribution of $\hat{\sigma}$, i.e., $\Pr(\sigma + \bar{w}_{A,\tau} \leq x)$, where $\bar{w}_{A,\tau} \sim \mathcal{N}(0, \sigma_{var}^2)$, and $\sigma_{var}^2 \triangleq \frac{1}{2\bar{P}L_{B,\tau}}$. Consider

$$G(x) \triangleq \Pr(\sigma + \bar{w}_{A,\tau} \leq x) \quad (30)$$

$$= \int_0^\infty f_\sigma(y) \int_{-\infty}^{x-y} \frac{1}{\sqrt{2\pi}\sigma_{var}} e^{-z^2/2\sigma_{var}^2} dz dy, \quad (31)$$

where $f_\sigma(y)$ is the pdf of σ , which is chi distributed with $2r$ degrees of freedom. Taking the derivative of (30) with respect to x , we get

$$\frac{\partial G(x)}{\partial x} = \frac{J}{\sqrt{2\pi}\sigma_{var}} \int_0^\infty y^{2r-1} e^{-\frac{y^2}{2}} e^{-\frac{(x-y)^2}{2\sigma_{var}^2}} dy \quad (32)$$

$$= \frac{J e^{-\beta_3}}{\sqrt{2\pi}\sigma_{var}} \int_0^\infty y^{2r-1} e^{\left\{-\frac{(y-\beta_1)^2}{2\beta_2}\right\}} dy, \quad (33)$$

where J is the constant term in the standard chi pdf, $\beta_1 \triangleq \frac{x}{1+\sigma_{var}^2}$, $\beta_2 \triangleq \frac{\sigma_{var}^2}{1+\sigma_{var}^2} \doteq \frac{1}{\bar{P}}$ and $\beta_3 \triangleq \beta_2 x^2 / (2\sigma_{var}^2)$. Let $t = \frac{y-\beta_1}{\sqrt{\beta_2}}$ and using the binomial expansion, it can be shown that

$$\begin{aligned} \frac{\partial G(x)}{\partial x} &= \frac{J \exp(-\beta_3)}{\sqrt{2\pi}\sigma_{var}} \sum_{j=0}^{2r-1} \binom{2r-1}{j} (\sqrt{\beta_2})^{2r-j} \\ &\times \frac{x^j}{(1+\sigma_{var}^2)^j} \int_{-\beta_1/\sqrt{\beta_2}}^\infty t^{2r-1-j} e^{-\frac{t^2}{2}} dt. \end{aligned} \quad (34)$$

Now, using $\exp(-\beta_3) \leq 1$, we can upper bound the first term in (29) as

$$\begin{aligned} \int_{\theta_P}^1 \frac{1}{x^{2s}} \frac{\partial G(x)}{\partial x} dx &\leq \frac{J}{\sqrt{2\pi}\sigma_{var}} \sum_{j=0}^{2r-1} \binom{2r-1}{j} C_j \\ &\times \frac{(\sqrt{\beta_2})^{2r-j}}{(1+\sigma_{var}^2)^j} \int_{\theta_P}^1 x^{j-2s} dx, \end{aligned} \quad (35)$$

where $s < r$, and $C_j \doteq 1$ is some constant that does not scale with \bar{P} . Now, the behavior of the terms above with \bar{P} is governed by

$$\frac{\beta_2^{r-j/2}}{\sigma_{var}} \int_{\theta_P}^1 x^{j-2s} dx \doteq \frac{1}{j-2s+1} \left[\frac{1}{\bar{P}^{a_1}} - \frac{1}{\bar{P}^{a_2}} \right], \quad (36)$$

where $a_1 \triangleq r - j/2 - 1/2$, and $a_2 \triangleq (-2s + j + 1)n + r - j/2 - 1/2$. The exponent corresponding to the first term above is $r - j/2 - 1/2 \geq 0$ for all $0 \leq j \leq 2r - 1$. Also, when $n = 1/2$, the exponent corresponding to the second term above is $r - s > 0$ for all $0 \leq j \leq 2r - 1$, and hence the integral is bounded for $1 \leq s < r$.

Finally, let $R_{\bar{P}} = g_m \log(\bar{P})$. Since $I_{\bar{P}} < \bar{P}$ and $F(\bar{P})$ are bounded when $0 \leq l \leq r + 1$, using $\left(\exp\left(\frac{L_c R_{\bar{P}}}{L_c - L_{B,\tau}}\right) - 1\right) \doteq \bar{P}^{\frac{g_m}{\alpha}}$ in (27), we get $\kappa_{\bar{P}} \doteq \frac{1}{\bar{P}^{\frac{g_m}{\alpha}-1}}$, where $\alpha \triangleq \frac{L_c - L_{B,\tau}}{L_c}$. This completes the proof of Lemma 1. ■

C. Proof of Theorem 1

Using the power control in (10), the outage probability in (8) can be written as

$$P_{out} = \Pr_{\{\hat{\sigma} \leq \theta_{\bar{P}}\}} \{\alpha \log(1 + \bar{P}^l \sigma^2) < R_{\bar{P}}\} \quad (37)$$

$$+ \Pr_{\{\hat{\sigma} > \theta_{\bar{P}}\}} \{\alpha \log(1 + \kappa_{\bar{P}} \Phi(\hat{\sigma}^{2s}) \sigma^2) < R_{\bar{P}}\} \quad (38)$$

$$\leq \Pi_1 + \Pi_2, \quad (39)$$

where $\Pi_1 \triangleq \Pr\{\alpha \log(1 + \bar{P}^l \sigma^2) < R_{\bar{P}}\}$, and $\Pi_2 \triangleq \Pr\{\alpha \log(1 + \kappa_{\bar{P}} \Phi(\hat{\sigma}^{2s}) \sigma^2) < R_{\bar{P}}\}$. In the above, we have used $\Pr_{\{A\}}\{\cdot\}$ to mean $\Pr\{\cdot \cap \{A\}\}$. Using $R_{\bar{P}} = g_m \log \bar{P}$, we have $\Pi_1 = \Pr\left\{\sigma^2 < \frac{1}{\bar{P}^{l-\frac{g_m}{\alpha}}}\right\}$ for large \bar{P} and $0 \leq l \leq r + 1$ from Lemma 1. From Lemma 3 in Appendix A, we have, $\Pi_1 \preceq \frac{1}{\bar{P}^{(l-\frac{g_m}{\alpha})r}}$. Next, substituting for $\Phi(\hat{\sigma}^{2s})$ from (9), Π_2 can be written as, $\Pi_2 = \Pr\{\sigma^2 < \hat{\sigma}^{2s} / \kappa_{\bar{P}}\}$. Using $\hat{\sigma}^2 \leq \hat{\sigma}_U^2 \triangleq (\sigma + |\bar{w}_{A,\tau}|)^2$ from Lemma 4 in Appendix A with $\hat{\sigma}^2 = \sigma^2$, we get

$$\Pi_2 \leq \Pr\left\{\sigma^2 < \frac{1}{\kappa_{\bar{P}}} (\sigma + |\bar{w}_{A,\tau}|)^{2s}\right\} \quad (40)$$

$$\begin{aligned} &\leq \Pr\left\{\sigma^2 < \frac{(2\sigma)^{2s}}{\kappa_{\bar{P}}} \cap \sigma^2 > |\bar{w}_{A,\tau}|^2\right\} \\ &+ \Pr\left\{\sigma^2 < \frac{(2|\bar{w}_{A,\tau}|)^{2s}}{\kappa_{\bar{P}}} \cap \sigma^2 \leq |\bar{w}_{A,\tau}|^2\right\}. \end{aligned} \quad (41)$$

It is straightforward to show that provided $\kappa_{\bar{P}}$ is strictly increasing with \bar{P} , the first term in the above goes to zero exponentially with \bar{P} for $1 \leq s < r$. This implies that $g_m < \alpha$, since $\kappa_{\bar{P}} \doteq \bar{P}^{(1-\frac{g_m}{\alpha})}$ from Lemma 1. The second term in (41) is upper-bounded as

$$\Pr\left\{\sigma^2 < \frac{|\bar{w}_{A,\tau}|^{2s} 2^{2s}}{\kappa_{\bar{P}}}\right\} \stackrel{(a)}{\leq} \frac{2^{2sr} \mathbb{E}|\bar{w}_{A,\tau}|^{2sr}}{\kappa_{\bar{P}}^r r!} \quad (42)$$

$$\stackrel{(b)}{=} \frac{1}{\bar{P}^{r(s+1-\frac{g_m}{\alpha})}}, \quad (43)$$

where (a) follows from Lemma 3 in Appendix A, and the \doteq in (b) uses the fact that $\kappa_{\bar{P}} \doteq \bar{P}^{(1-\frac{g_m}{\alpha})}$ and $\mathbb{E}|\bar{w}_{A,\tau}|^{2sr} \doteq 1/\bar{P}^{sr}$. Hence, we have $\Pr\left\{\sigma^2 < \frac{|\bar{w}_{A,\tau}|^{2s} 2^{2s}}{\kappa_{\bar{P}}}\right\} \preceq \frac{1}{\bar{P}^{r(s+1-\frac{g_m}{\alpha})}}$, which implies $\Pi_2 \preceq \frac{1}{\bar{P}^{r(s+1-\frac{g_m}{\alpha})}}$. Using this and $\Pi_1 \preceq \frac{1}{\bar{P}^{r(l-\frac{g_m}{\alpha})}}$ in (39), we have

$$P_{out} \preceq \max\left(\frac{1}{\bar{P}^{r(l-\frac{g_m}{\alpha})}}, \frac{1}{\bar{P}^{r(s+1-\frac{g_m}{\alpha})}}\right) \quad (44)$$

$$= \frac{1}{\bar{P}^{r(\min\{l, s+1\} - \frac{g_m}{\alpha})}}, \quad (45)$$

for $0 \leq l \leq r + 1$, $1 \leq s < r$ and $0 \leq g_m < \alpha$. This ends the proof of Theorem 1. ■

D. Proof of Lemma 2

Note that $\tilde{\mathbf{p}}_c$ can be written as

$$\tilde{\mathbf{p}}_c = \mathbf{p}_c - \tilde{\mathbf{y}}_{B,\tau_2} - \mathbb{E}\{\mathbf{p}_c - \tilde{\mathbf{y}}_{B,\tau_2} | \tilde{\mathbf{y}}_{B,\tau_2}\} \quad (46)$$

$$= \frac{1}{\sqrt{P}L_{A,\tau_2}} [\mathbb{E}\{\mathbf{w}_{B,\tau_2} | \tilde{\mathbf{y}}_{B,\tau_2}\} - \mathbf{w}_{B,\tau_2}]. \quad (47)$$

Now,

$$\mathbb{E}\|\tilde{\mathbf{p}}_c\|_2^{2z} = \frac{1}{\bar{P}^z L_{A,\tau_2}^z} \mathbb{E}_{\mathbf{w}_{B,\tau_2}, \tilde{\mathbf{y}}_{B,\tau_2}} \{\mathcal{A}\} \quad (48)$$

$$\stackrel{(a)}{\leq} \frac{1}{\bar{P}^z L_{A,\tau_2}^z} [\mathbb{E}_{\tilde{\mathbf{y}}_{B,\tau_2}} \{2^{2z} \|\mathbb{E}\{\mathbf{w}_{B,\tau_2} | \tilde{\mathbf{y}}_{B,\tau_2}\}\|_2^{2z}\} + 2^{2z} \mathbb{E}_{\mathbf{w}_{B,\tau_2}} \{\|\mathbf{w}_{B,\tau_2}\|_2^{2z}\}] \quad (49)$$

$$\stackrel{(b)}{\leq} \frac{2^{2z+1}}{\bar{P}^z L_{A,\tau_2}^z} \mathbb{E}\|\mathbf{w}_{B,\tau_2}\|_2^{2z} \doteq \frac{1}{\bar{P}^z}, \quad (50)$$

where $\mathcal{A} \triangleq \|\mathbb{E}\{\mathbf{w}_{B,\tau_2} | \tilde{\mathbf{y}}_{B,\tau_2}\} - \mathbf{w}_{B,\tau_2}\|_2^{2z}$. In the above, (a) follows from the triangle inequality and using $(a+b)^n \leq (2a)^n + (2b)^n$ for even $n > 0$, and (b) follows from the Jensen's inequality and the fact that $\mathbb{E}\|\mathbf{w}_{B,\tau_2}\|_2^{2z} < \infty$ as $\bar{P} \rightarrow \infty$. The subscripts on the expectation in the above denote the random variables over which the expectation is taken; and $\mathbb{E}\{X|y\}$ denotes the expectation of the random variable X conditioned on the instantiation $Y = y$. This completes the proof. ■

E. Proof of Theorem 2

Using the capacity lower bound in (17), the outage probability can be upper bounded as

$$P_{out} \leq \Pr\{C_{AB} < R_{\bar{P}}\}, \quad (51)$$

where $R_{\bar{P}} \triangleq g_m \log \bar{P}$ is the target data rate. We choose $\eta < 1$, and arbitrarily close to 1. We split the event in the expression for P_{out} as

$$\begin{aligned} P_{out} &\leq \Pr\left\{C_{AB} < R_{\bar{P}} \cap \mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B,\tau_2}] \leq \frac{1}{\bar{P}^\eta}\right\} \\ &\quad + \Pr\left\{C_{AB} < R_{\bar{P}} \cap \mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B,\tau_2}] > \frac{1}{\bar{P}^\eta}\right\} \\ &\stackrel{(a)}{\leq} \Pr\left\{\alpha \log\left(1 + \frac{\bar{P}\|\hat{\mathbf{p}}_c\|_2^2}{\bar{P}^{(1-\eta)} + 1}\right) < R_{\bar{P}}\right\} \\ &\quad + \Pr\left\{\mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B,\tau_2}] > \frac{1}{\bar{P}^\eta}\right\}, \end{aligned} \quad (53)$$

where (a) follows by substituting $1/\bar{P}^\eta$ in place of $\mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B,\tau_2}]$ in the first term, and removing one of the events in the intersection. Define $\bar{R}_{\bar{P}} \triangleq \frac{(\exp\{R_{\bar{P}}/\alpha\}-1)}{\bar{P}} \left(\frac{\bar{P}^{(1-\eta)}}{r} + 1\right)$, and note that $\bar{R}_{\bar{P}} \doteq \frac{1}{\bar{P}^{(\eta-\frac{g_m}{\alpha})}}$. Then, the first term in (53) can be written as:

$$\begin{aligned} \Pr\{\|\hat{\mathbf{p}}_c\|_2^2 < \bar{R}_{\bar{P}}\} &\stackrel{(a)}{\leq} \Pr\left\{\|\mathbf{p}_c\|_2 - \|\tilde{\mathbf{p}}_c\|_2 < \sqrt{\bar{R}_{\bar{P}}}\right\} \\ &\leq \Pr\{E_1 \cap E_2\} + \Pr\{E_1 \cap E_2^c\} \\ &\leq \Pr\left\{\|\tilde{\mathbf{p}}_c\|_2 > \sqrt{\bar{R}_{\bar{P}}}\right\} \\ &\quad + \Pr\{\|\mathbf{p}_c\|_2^2 < 4\bar{R}_{\bar{P}}\}, \end{aligned} \quad (55)$$

where $E_1 \triangleq \{\|\mathbf{p}_c\|_2 < \|\tilde{\mathbf{p}}_c\|_2 + \sqrt{\bar{R}_{\bar{P}}}\}$ and $E_2 \triangleq \{\|\tilde{\mathbf{p}}_c\|_2 > \sqrt{\bar{R}_{\bar{P}}}\}$. In the above, (a) follows from the reverse triangle inequality, and the last two inequalities follow by ignoring

one of the events in the intersection. The first term in (55) can be written as

$$\Pr\{\|\tilde{\mathbf{p}}_c\|_2^{2\delta} > \bar{R}_{\bar{P}}^\delta\} \stackrel{(a)}{\leq} \frac{\mathbb{E}\|\tilde{\mathbf{p}}_c\|_2^{2\delta}}{\bar{R}_{\bar{P}}^\delta} \stackrel{(b)}{\leq} \frac{1}{\bar{P}^\delta} \frac{1}{\bar{P}^{(\frac{g_m}{\alpha}-\eta)\delta}}, \quad (56)$$

where (a) follows from the Markov inequality and (b) follows from Lemma 2. Letting $\delta = r \frac{1}{\frac{g_m}{\alpha}-\eta+1} (s+1 - \frac{g_m}{\alpha}) > 0$, we have

$$\Pr\left\{\|\tilde{\mathbf{p}}_c\|_2 > \sqrt{\bar{R}_{\bar{P}}}\right\} \leq \frac{1}{\bar{P}^{r(s+1-\frac{g_m}{\alpha})}}, \quad 1 \leq s < r. \quad (57)$$

In order to solve for the second term in (55), we need to handle two cases of the singular value estimate at node A separately; the good estimated channel case $g \triangleq \{\hat{\sigma} \geq \theta_{\bar{P}}\}$ and the bad estimated channel case $b \triangleq \{\hat{\sigma} < \theta_{\bar{P}}\}$.

1) *Good Estimated Channel* $\{\hat{\sigma} \geq \theta_{\bar{P}}\}$: When $\hat{\sigma} \geq \theta_{\bar{P}}$, substituting for $\mathbf{p}_c \triangleq \sqrt{\bar{P}(\hat{\sigma})}\mathbf{h}$ and $\kappa_{\bar{P}} \doteq \bar{P}^{-\frac{g_m}{\alpha}}$, and defining $\hat{\sigma}_U \triangleq (\sigma + |\bar{w}_{A,\tau}|)$ as the upper bound on $\hat{\sigma}$ from Lemma 4 in Appendix A, the second term in (55) leads to:

$$\begin{aligned} \Pr_{\{\hat{\sigma} \geq \theta_{\bar{P}}\}}\{E_3\} &\stackrel{(a)}{\leq} \Pr\left\{\sigma^2 < 2^{2(s+1)} \sigma^{2s} \bar{R}_{\bar{P}} \cap E_4\right\} \\ &\quad + \Pr\left\{\sigma^2 < 2^{2(s+1)} |\bar{w}_{A,\tau}|^{2s} \bar{R}_{\bar{P}} \cap E_4\right\} \\ &\leq \Pr\left\{\sigma^{2(s-1)} > \frac{2^{-2(s+1)}}{\bar{R}_{\bar{P}}}\right\} \\ &\quad + \Pr\left\{\sigma^2 < 2^{2(s+1)} |\bar{w}_{A,\tau}|^{2s} \bar{R}_{\bar{P}}\right\}, \end{aligned} \quad (58)$$

where $E_3 \triangleq \{\frac{\|\mathbf{h}\|_2^2}{\hat{\sigma}_U^{2s}} < 4\bar{R}_{\bar{P}}\}$, and $E_4 \triangleq \{\sigma^2 > |\bar{w}_{A,\tau}|^2\}$.

In the above, we have used $\Pr_{\{A\}}\{\cdot\}$ to mean $\Pr\{\cdot \cap \{A\}\}$, as before; and (a) follows by ignoring the event g . It can be shown that first term in (58) decreases exponentially with $\bar{P}^{\frac{\eta-g_m/\alpha}{s-1}}$, as follows:

$$\begin{aligned} \Pr\{\mathcal{B}\} &\stackrel{(a)}{=} \int_{1/\bar{R}_{\bar{P}}^{1/(s-1)}}^\infty e^{-x} x^{r-1} dx \\ &\doteq \exp\left\{-1/\bar{R}_{\bar{P}}^{1/(s-1)}\right\} \sum_{k=0}^{r-1} \frac{1}{(\bar{R}_{\bar{P}}^{1/(s-1)})^{r-k-1}} \\ &\stackrel{(b)}{=} e^{-\mathcal{Z}}, \end{aligned} \quad (60)$$

where $\mathcal{B} \triangleq \{\sigma^{2(s-1)} > \frac{1}{2^{2(s+1)} \bar{R}_{\bar{P}}}\}$, and $\mathcal{Z} \triangleq \bar{P}^{\frac{\eta-g_m/\alpha}{s-1}}$. In the above, (a) follows by ignoring the constant factors and substituting for the chi-square pdf of σ^2 . Since $1/\bar{R}_{\bar{P}} \doteq \bar{P}^{(\eta-g_m/\alpha)}$ when $g_m < \eta\alpha$, and since the exponential term outside the summation dominates the polynomial terms inside the summation, we obtain (b). Note that the special case of $s = 1$ is trivial, since this corresponds to the probability that $\bar{R}_{\bar{P}}$ exceeds a constant, which becomes 0 for sufficiently large \bar{P} . The second term in (58) becomes:

$$\begin{aligned} \Pr\left\{\sigma^2 < 2^{2(s+1)} |\bar{w}_{A,\tau}|^{2s} \bar{R}_{\bar{P}}\right\} &\leq \frac{2^{2r(s+1)} \bar{R}_{\bar{P}}^r \mathbb{E}\{|\bar{w}_{A,\tau}|^{2sr}\}}{r!} \\ &\doteq \frac{1}{\bar{P}^{(\eta-\frac{g_m}{\alpha})r} \bar{P}^{rs}} \end{aligned} \quad (61)$$

$$= \frac{1}{\bar{P}^{r(s+\eta-g_m/\alpha)}} \quad (62)$$

for $0 \leq g_m < \eta\alpha$. In the above, we have used Lemma 3 in Appendix A and $\mathbb{E}|\bar{w}_{A,\tau}|^{2s} = \frac{1}{P^s}$. Thus, in the good estimated channel case, we have

$$\Pr_{\{\hat{\sigma} \geq \theta_{\bar{P}}\}} \left\{ \|\mathbf{p}_c\|_2^2 \leq 4\bar{R}_{\bar{P}} \right\} \preceq \frac{1}{\bar{P}^{r(s+\eta-\frac{gm}{\alpha})}}, \quad 0 \leq g_m < \eta\alpha. \quad (63)$$

2) *Bad Estimated Channel* $\{\hat{\sigma} < \theta_{\bar{P}}\}$: Recall that when $\hat{\sigma} < \theta_{\bar{P}}$, the composite channel is given by $\mathbf{p}_c = \sqrt{\bar{P}l}\mathbf{h}$. With this, the second term in (55) becomes

$$\Pr_{\{\hat{\sigma} < \theta_{\bar{P}}\}} \left\{ \|\mathbf{p}_c\|_2^2 < 4\bar{R}_{\bar{P}} \right\} = \Pr \left\{ \|\mathbf{h}\|_2^2 < \frac{4\bar{R}_{\bar{P}}}{\bar{P}l} \cap b \right\} \leq \Pr \left\{ \sigma^2 < \frac{4\bar{R}_{\bar{P}}}{\bar{P}l} \right\} \quad (64)$$

$$\doteq \frac{1}{\bar{P}^{rl}} \frac{1}{\bar{P}^{r(-\frac{gm}{\alpha} + \eta)}} \quad (65)$$

$$\doteq \frac{1}{\bar{P}^{r(l+\eta-\frac{gm}{\alpha})}}, \quad (66)$$

where $0 \leq l \leq r$. This completes the analysis of the first term in (53).

Now, the second term in (53) can be bounded as:

$$\Pr \left\{ (\mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B,\tau_2}])^\zeta > \frac{1}{\bar{P}^\zeta \eta} \right\} \stackrel{(a)}{\leq} \mathbb{E}(\mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B,\tau_2}])^\zeta \bar{P}^\zeta \eta \stackrel{(b)}{\leq} \mathbb{E}([\|\tilde{\mathbf{p}}_c\|_2^{2\zeta}]) \bar{P}^\zeta \eta \quad (67)$$

$$\stackrel{(c)}{\preceq} \frac{1}{\bar{P}^\zeta (1-\eta)}, \quad (68)$$

where $\zeta > 0$ is an arbitrary number. In the above, (a) and (b) follow from the Markov inequality and Jensen's inequality, respectively, and (c) follows from Lemma 2. Since $\eta < 1$, and ζ can be chosen arbitrarily large, the second term in (53) goes to zero with an arbitrarily large exponent as \bar{P} goes to infinity.

Putting (57), (63), (66) and (68) together, a DMT of $d(g_m) = r(\min\{l, s\} + \eta - \frac{gm}{\alpha})$ is achievable, for $0 \leq l \leq r$, $1 \leq s < r$ and $0 \leq g_m < \eta\alpha$. Noting that η is arbitrarily close to 1 completes the proof of Theorem 2. ■

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers for their detailed comments and suggestions, which have improved the quality of this paper.

REFERENCES

- [1] B. Bharath and C. Murthy, "On the improvement of diversity-multiplexing gain tradeoff in a training based TDD-SIMO system," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Sig. Proc. (ICASSP)*, 2010, pp. 3366–3369.
- [2] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
- [3] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1273–1289, Apr. 2001.
- [4] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, May 1999.
- [5] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [6] H. El Gamal, G. Caire, and M. Damen, "The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3601–3621, Aug. 2006.
- [7] A. Khoshnevis and A. Sabharwal, "On diversity and multiplexing gain of multiple antenna systems with transmitter channel information," in *Proc. Allerton Conf. on Commun., Control and Computing*, 2004.
- [8] H. N. Raghava and V. Sharma, "Diversity-multiplexing trade-off for channels with feedback," in *Allerton Conf. Commun., Control, and Computing*, 2005, pp. 668–677.
- [9] L. Zheng, "Diversity-multiplexing tradeoff: A comprehensive view of multiple antenna systems," Ph.D. dissertation, University of California, Berkeley, 2002.
- [10] A. Lim and V. Lau, "On the fundamental tradeoff of spatial diversity and spatial multiplexing of MIMO links with imperfect CSIT," in *IEEE Int. Symp. on Inf. Theory*, 2006, pp. 2704–2708.
- [11] —, "On the fundamental tradeoff of spatial diversity and spatial multiplexing of MISO/SIMO links with imperfect CSIT," *IEEE Trans. on Wireless Commun.*, vol. 7, no. 1, pp. 110–117, Jul. 2008.
- [12] T. Kim and G. Caire, "Diversity gains of power control with noisy CSIT in MIMO channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1618–1626, Apr. 2009.
- [13] C. Steger and A. Sabharwal, "Single-input two-way SIMO channel: Diversity-multiplexing tradeoff with two-way training," *IEEE Trans. on Wireless Commun.*, vol. 7, no. 12, pp. 4877–4885, Dec. 2008.
- [14] X. J. Zhang, Y. Gong, and K. B. Letaief, "On the diversity gain in cooperative relaying channels with imperfect CSIT," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1273–1279, Apr. 2010.
- [15] A. Khoshnevis and A. Sabharwal, "Performance of quantized power control in multiple antenna systems," in *IEEE International Conf. on Commun.*, vol. 2, 2004, pp. 803–807.
- [16] T. Kim and M. Skoglund, "Diversity-multiplexing tradeoff in MIMO channels with partial CSIT," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2743–2759, Aug. 2007.
- [17] V. Aggarwal and A. Sabharwal, "Bits about the channel: multiround protocols for two-way fading channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3352–3370, Jun. 2011.
- [18] X. J. Zhang, Y. Gong, and K. B. Letaief, "On the diversity gain in MIMO channels with joint rate and power control based on noisy CSIT," *IEEE Trans. on Wireless Commun.*, vol. 10, no. 1, pp. 68–72, Jan. 2011.
- [19] —, "Power control and channel training for MIMO channels: A DMT perspective," *IEEE Trans. on Wireless Commun.*, vol. 10, no. 7, pp. 2080–2089, Jul. 2011.
- [20] B. N. Bharath and C. R. Murthy, "Reverse channel training for reciprocal MIMO systems with spatial multiplexing," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, Taipei, Taiwan, Apr. 2009, pp. 2673–2676.
- [21] X. Zhou, T. Lamahewa, P. Sadeghi, and S. Durrani, "Two-way training: optimal power allocation for pilot and data transmission," *IEEE Trans. on Wireless Commun.*, vol. 9, no. 2, pp. 564–569, Feb. 2010.
- [22] B. N. Bharath and C. R. Murthy, "Power controlled reverse channel training achieves an infinite diversity order with perfect CSIR," *Under Preparation*, 2011.
- [23] S. Tavildar and P. Viswanath, "Approximately universal codes over slow-fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3233–3258, Jul. 2006.
- [24] P. Elia, K. R. Kumar, S. A. Pawar, P. V. Kumar, and H.-F. Lu, "Explicit space-time codes achieving the diversity-multiplexing gain tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3869–3884, Sep. 2006.
- [25] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.
- [26] C. Steger, A. Khoshnevis, A. Sabharwal, and B. Aazhang, "The case for transmitter training," in *IEEE Int. Symp. on Inf. Theory*, Washington, U.S.A., Sep. 2007.
- [27] D. Guo, Y. Wu, S. Shamai, and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.
- [28] Y. Liang and V. Veeravalli, "Capacity of noncoherent time-selective rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095–3110, 2004.